Project Description

1 Background and Overview

This project pursues a new approach to the development of theories of phonological grammar and learning, with the goal of shedding new light on an empirical domain that poses challenges for previous theories: phonological opacity. We use this term to refer to the set of phenomena that can be analyzed in derivational terms as generalizations that hold of an intermediate level of representation, but not of the observed surface level. Canadian English diphthongs provide a famous example. The raised variants [ʌɪ]/[ʌʊ] of [aɪ]/[aʊ] appear before voiceless obstruents (e.g. [ɹʌɪ] ‘write’ vs. [ɹaɪ] ‘ride’ and [ɹaʊ] ‘rye’). The generalization that raising only occurs in this environment is not true at the surface, however, because there are words with raising before a voiced flap (e.g. [mʌɪɾɚ] ‘mitre’ vs. [sʌɪɾɚ] ‘cider’). In the standard derivational analysis [1], the “raised only before voiceless” generalization is true of an intermediate level of representation, in which the flaps that condition raising are voiceless stops (i.e. /mʌɪɾɚ/ → mʌɪɾɚ → [mʌɪɾɚ] vs. /sʌɪɾɚ/ → sʌɪɾɚ → [sʌɪɾɚ]).

Our study of phonological opacity will involve evaluating novel grammatical theories simultaneously for both their typological fitness and their learnability. The standard approach to theory development in phonology is sequential: first construct a theory of grammar that comes as close as possible to generating all and only the cross-linguistically attested types of phonological systems, and then work toward a formal learning theory that will arrive at a correct grammar for any language within the space defined by the grammatical theory. This sequential approach is the basis of the theories of phonological grammar and learning that have been developed in the original version of generative phonology [2, 3], as well as in subsequent work in the principles and parameters framework [4, 5], and in Optimality Theory (OT) [6, 7].

Due to recent developments in phonological theory and in computational approaches to learning, some of which were the outcome of our last NSF-sponsored project, it is now feasible to assess novel theories of phonological grammar in terms of how they contribute to a learning theory, even as those theories are being developed. We expect that this new approach will speed the development of theories of phonology that are both typologically supported and learnable. We also expect that it will contribute to the development of new theories of learning and to new hypotheses about human language acquisition.

Our research will focus on grammatical theories that combine constraint-based and derivational formalisms. Constraint-based frameworks, in particular OT, have proven particularly suitable for the development of theories with a good fit to attested typology [8]. OT has also supported the development of relatively successful theories of learning [7, 9]. Since its inception, however, phonological opacity has posed a significant challenge to OT [10, 11]. Because the original version of OT has no levels of representation intermediate between underlying and surface (OT’s input and output), it cannot replicate the standard derivational analyses, and very few cases of opacity can be analyzed with just rankings of standard OT’s markedness and faithfulness constraints [12]. A variety of solutions to this problem have been proposed, either using only standard OT’s two levels but enriching its constraint types or representations [13-18], or else pursuing revised versions of OT that incorporate richer derivations with intermediate levels [19, 20]. Our main focus will be on developing and comparing such derivational versions of OT.

Crucially, combining constraints with richer derivations does not automatically lead to a theory that generates a superset of the languages generated by classic OT. In the last grant period, we conducted an assessment of the overall typological predictions of one derivational
version of OT, Harmonic Serialism (HS). We found a number of ways in which the predictions of HS are in fact more restrictive than classic OT, and which result in a better match to the attested cross-linguistic typology. In our view, the main challenge facing HS is that, like standard OT, it does not provide a very general account of opacity. For example, it cannot replicate the derivational analysis of the interaction between Canadian raising and flapping. One of the primary goals of this project is to develop a version of HS that retains its desirable qualities in terms of restrictiveness, but which can cope with a wider range of opaque phenomena. This will require a departure from the theory of opacity developed by co-PI McCarthy prior to the last grant period, OT with Candidate Chains [20], since that theory overgenerates in several undesirable ways. We also seek an alternative to this theory on learning grounds, as we discuss in Sections 6 and 7.

The data on opacity that we will consider in developing this theory will come from a new cross-linguistic investigation, which will also lead to the creation of a new publicly accessible database. Opacity has been much studied in phonological theory, especially in the years immediately following the publication of the Sound Pattern of English [21] in 1968. We will aim to collect at least all of the cases of opacity mentioned thus far in the theoretical literature, and will update the descriptions using the most recent primary sources on the languages. We will make our collection process public, so that other linguists will be able to contribute further cases. Unlike the descriptions one might find in problem sets, and even sometimes in the theoretical literature, the database will remain as faithful as possible to the original sources in retaining any references to exceptionality, (non-)productivity, variation, or morpho-syntactic conditioning in the application of the processes interacting in an opaque fashion.

Opacity poses challenges not only for theories of grammar, but also for theories of learning. Alongside its typological successes, one of the strengths of classic OT is that it has learning algorithms that succeed in finding a correct constraint ranking for a given set of data, if such a ranking exists [7, 22]. The standard OT learning algorithms are not guaranteed to succeed, however, if they are not provided with the full structure of the learning data, including underlying representations and prosodic structure. The intermediate levels of a derivational account of opacity create further instances of such potentially problematic ‘hidden structure’.

The OT learning algorithms are not only of theoretical interest, but also of considerable practical utility. They have been implemented in freely available software programs, which aid in the analysis of individual languages, and in the generation of typological predictions [23-25]. One of the main successes of our last project was the development of an algorithm for the generation of typological predictions for HS, and its dissemination in the freely available OT-Help [26]. As we develop extensions of HS that better cope with opacity, we will also develop extensions of OT-Help that allow us and other researchers to assess the typological predictions of these theories, as well as related alternatives.

We also developed a general method for learning of HS derivations [27], which copes both with hidden structure and with variation in the output for a given input. Like most methods for hidden structure learning developed for classic OT, as well as methods developed for dealing with structural ambiguity in natural language processing, our method is not guaranteed to succeed on all cases. It has been successful on the small set of cases we have examined thus far, but it remains to be more rigorously tested on a set of patterns representative of those found cross-linguistically. Alongside the development of accounts of opacity, and the opacity database, the further development and testing of our model for learning of derivations and other hidden structure is the primary focus of our planned research. Models of hidden structure learning in OT
have thus far been compared on a set of word stress learning problems developed in the original OT hidden structure research [7]. We will develop an improved and expanded set of test cases that will guide our learning research, and which we will make freely available to others.

Our approach to hidden structure learning builds on applications of Maximum Entropy (MaxEnt) models [28] to phonological learning [29-31]. These models were originally adopted because of their usefulness in dealing with phonological variation. Unlike earlier learning algorithms that were developed specifically for OT, the MaxEnt models used for phonological learning have been broadly applied elsewhere, especially in natural language processing. MaxEnt models thus have the considerable advantages that their mathematical properties are well understood, and that there are existing methods for dealing with structural ambiguity.

This strong connection between our learning research and research in computer science forms the basis for the primary broader impact of our project. We aim to help to increase the representation of women in science by providing training in the use of computational methods in linguistics research, and access to further training in computer science. The planned graduate student RAs for our project are all women with some background and interest in programming and mathematics. We will also seek linguistic undergraduates to participate in an undergraduate mentoring program, in which they will be directly supervised by our graduate RAs, thus providing examples of women engaged in computational research, and of the usefulness of computational methods in furthering our understanding of language.

The rest of this proposal will proceed by first discussing the results of our previous support and their relevance to our planned research, and then moving on to the further details of the grammatical and learning work, including database and software development. Section 7 provides a schedule for the planned research as well as assessment and management plans, and section 8 focuses on broader impact.

2 Previous Support

2.1 Intellectual merit of previous work
The results of this grant include OT-Help 2.0 [26], a software environment for studying Harmonic Serialism and Harmonic Grammar (HG), implementations of learning algorithms in the popular statistical software environment R [32], 14 articles published or accepted in refereed journals [17, 33-45], an edited volume under contract [46], three doctoral dissertations [47-49], and a great many book chapters, conference papers, posters, and other works (e.g. [27, 50-70], among others.).

As we explained briefly in the last section, HS is a version of OT with richer derivations; we will discuss the structure of this theory in more depth in section 3. HG is a version of OT in which ranked constraints are replaced by weighted ones [41, 71]; constraint weighting will be discussed in relation to our learning work in section 4. The main aim of our last grant was to investigate whether these two changes to the standard version of OT, separately or jointly, would lead to improved typological predictions. To aid this investigation, we developed OT-Help 2.0 [26], which extends the capabilities of OT-Soft [23] in two ways. First, we adopted OT-Soft’s approach to OT typology calculation using Recursive Constraint Demotion [7], and extended it by adding an application of Linear Programming’s simplex algorithm [35], which allows for the calculation of typologies with weighted rather than ranked constraints. Second, we created a typology calculator for serial versions of OT (i.e., HS) and HG (Serial HG), which involved implementing functions that allow users to define their own candidate-creating operations as well
as the constraints that evaluate those candidates. These operations and constraints can also be used to generate candidates and violations for some cases in standard parallel OT. They have been used for this purpose in comparing serial and parallel predictions in stress typology [49], and we will also use them in generating candidate sets for some of the learning experiments described in section 6. (A more general parallel Gen is implemented in [72].) The broad usefulness of this software is enhanced by its being a platform-independent freely downloadable Java application.

In addition to OT-Help 2.0, the results of the previous grant can be grouped into four broad categories. The results in (i), (ii), and (iv) are most relevant to this new proposal, and they will be discussed in greater detail below, as they become relevant.

i. Various works written under the earlier grant [17, 37, 38, 54, 60-62, 64, 68, 73, 74] have explored how serial models of constraint satisfaction differ from parallel models. Particularly relevant to the new project are the findings about the analysis of opacity in serial models.

ii. Other works [35, 36, 41, 42, 47, 48, 56, 57, 59, 67-69, 75] identify a range of ways in which weighted constraint grammars differ from (and are arguably superior to) ranked constraint grammars, and show that the predictions of HG are improved when combined with a serial model, Serial HG.

iii. Still other research [44, 51-53, 70] has implications for the structure of the lexicon and the analysis of irregular processes and exceptions.

iv. Several papers [27, 36, 47, 55, 76] investigate the nature of learning in serial theories and Harmonic Grammar.

2.2 Broader impact of previous work
A total of 12 doctoral students, 7 of whom are women, received financial support from this grant. Four have completed their degrees and gone on to faculty or post-doctoral positions at USC, Manchester, McGill and Arizona State. The other 8 are making good progress toward their degrees. The work supported by our last grant also had an educational impact beyond training our own graduate students. We have had reports that the research supported by our grant is being read in many graduate seminars, and that OT-Help 2.0 is being used in teaching and research in other locations both in the United States and abroad. A course “Harmonic Grammar: Models and Methods” was offered by Pater at the LSA summer institute in Boulder Colorado in the spring of 2011; it introduced students both to some of the results of our research and to OT-Help 2.0 and the R-implemented learning software. The materials from the course are available for others’ use on the internet. We were successful at building interdisciplinary bridges in our grant-supported research (see esp. [41]). We also played a central role in the establishment of the Institute for Computational and Experimental Study of Language at our university (see further section 8), which has a goal of supporting interdisciplinary training like that of our last grant project and of the present proposal. Finally, we were successful in recruiting members of underrepresented groups into our associated REU program.

3 Planned grammatical research: Opacity in Harmonic Serialism
There are two main differences between Harmonic Serialism (HS) ([6, 8, 26, 33, 34, 37-40, 46, 48-54, 57, 59-65, 68, 69, 73, 74, 77-83]; see also [84-87] for related approaches) and the standard parallel model of OT ([6] and others). First, in HS the candidate-generating component GEN is limited to making one change at a time, so the candidates produced by GEN differ only
minimally from its input. In parallel OT, Gen can make many changes at once, in parallel. Second, in HS the output of Eval is submitted as a new input to Gen, a process that continues until convergence, when the most recent input to Gen and the output of Eval are identical. In strictly parallel OT, the surface form is reached after a single pass through Gen and Eval. (Cyclic and Stratal versions of OT allow several passes, but only in conjunction with the morphology.)

Extant arguments in support of HS fall into two types, both of which were explored in our previous work and will remain important in this project. One type of argument is based on the fact that HS allows intermediate derivational steps while parallel OT does not. Generalizations can be non-surface-true or, indeed, unintelligible in surface forms because they rely on structure that was present at an intermediate step and is later changed. An example is the interaction of stress and syncope [73]. In many languages, unstressed vowels delete. This seemingly straightforward phenomenon proves surprisingly difficult to analyze in parallel OT [88, 89], though, because the generalization is inherently serial in character: stress is assigned and then the unstressed vowels are targeted for deletion. In HS, the serial character of this generalization is expressed directly: an OT grammar is ranked so as to assign stress and then to delete the unstressed vowels, in successive steps of the derivation. Other arguments for intermediate steps in HS can be found in [34, 37, 38, 51, 57, 60, 68, 69, 83].

The other form of argument for HS is typological. OT is inherently typological: given a constraint set, a language typology emerges from ranking permutation. Even with identical constraint sets, the typologies of HS and parallel OT are different. For example, simultaneous satisfaction of constraints on stress and constraints on unstressed vowels predicts unattested patterns of stress-syncope interaction that serial satisfaction does not predict [73]. This shows that assignment of stress and deletion of vowels cannot be done in parallel — they must be done serially, as HS requires. Other typological arguments for HS can be found in [48-50, 59, 62, 64, 65, 74, 79, 82]. A typological argument against HS is presented in [90], but see [33] for a rejoinder.

According to Kiparsky’s classic definition, a phonological process P is opaque if there are surface forms that look like they should have undergone P but have not, or if there are surface forms that have undergone P but look like they have not [91]. More generally, P is opaque if the reason why it did not apply, the reason why it did apply, or how it applied are unexplained in surface representation. Stress-syncope interaction is a type of phonological opacity: the generalization about which vowels are deleted cannot be stated on surface representation (nor, indeed, on underlying representation either).

In rule-based phonology, rule ordering is the primary means of controlling rule interaction. The rule orderings that produce opacity are referred to as counterfeeding and counterbleeding [91]. In counterfeeding opacity, a rule ordered after P creates surface forms that look like they should have undergone P but have not, because P has already applied. In counterbleeding opacity, P applies and a rule ordered later alters its effects or the conditions that made it applicable.

Unmodified HS is capable of analyzing some types of counterbleeding opacity, using constraint ranking to produce effects similar to rule ordering. Stress-syncope interaction is one

---

1 Although this description could equally well apply in situations where P has lexical exceptions, the term opacity is traditionally reserved for those cases where P is opaque because of how it interacts with other processes in the language. We adopt this more restrictive definition because these are the cases that are challenging to OT.
Another example is stress-epenthesis interaction in some varieties of Arabic. An otherwise quite regular process assigning stress to heavy penults has surface exceptions when the penult contains an epenthetic vowel: /katab-l-ha/ → [kaˈtabilha] ‘he wrote to her’. In the standard rule-based analysis, stress is assigned to the penult before epenthesis applies [92-94]: /katab-l-ha/ → kaˈtabilha → [kaˈtabilha]. The generalization that heavy penults are stressed is true of the intermediate stage before the application of epenthesis. In an HS analysis [60], the intermediate step is produced by ranking the constraint against stressless words (HEAD(word)), which forces stress assignment, above the constraint against unsyllabified consonants (PARSE(seg)), which forces epenthesis. Because the HS Gen is limited to making a single change, the candidate set in the first pass includes a candidate with stress added, and one with epenthetic vowel, but not one with both changes (unsyllabified consonants are italicized):

<table>
<thead>
<tr>
<th>First pass: stress assignment</th>
<th>/katab-l-ha/</th>
<th>HEAD(word)</th>
<th>PARSE(seg)</th>
<th>DEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. → kaˈtabilha</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. katabilha</td>
<td>*!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. katabilha</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the second pass, epenthesis applies due to the ranking of PARSE(seg) over DEP:

<table>
<thead>
<tr>
<th>Second pass: epenthesis</th>
<th>kaˈtabilha</th>
<th>HEAD(word)</th>
<th>PARSE(seg)</th>
<th>DEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. → kaˈtabilha</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. kaˈtabilha</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this approach, the stress process is not ordered before the epenthesis process. Instead, the markedness constraints that compel stress and epenthesis determine which change happens first. The stress-syncope interaction discussed above is similarly opaque, and the HS analysis depends on a similar constraint ranking.

Although analyses like these look promising, unmodified HS is not a full theory of opacity. There are some examples of counterbleeding opacity that cannot be analyzed in HS, and HS is not in general applicable to counterfeeding interactions [8, 20, 63, 78]. Two research questions emerge from this. One question is empirical: of the well-understood, productive cases of counterbleeding opacity, which can and cannot be analyzed in HS, and is there a principled difference between the two? The other question is theoretical: what is the best way to extend HS so it can cope with counterfeeding opacity and the residue of intractable examples of counterbleeding opacity?

4 Opacity Database

To answer these empirical and theoretical questions, our first step will be to conduct an exhaustive survey of the known cases of phonological opacity in the theoretical literature, and to update their descriptions on the basis of all currently available primary sources on the languages. We will categorize the patterns in terms of their rule-based analyses: whether they are instances of counterfeeding or counterbleeding opacity, further subdivided as to whether opacity arises as a result of a change to the focus of the earlier applying rule, or to its environment. These categories
have proven to be useful in understanding the successes, and the limitations, of earlier constraint-based approaches to opacity, which typically deal with only a subset of them (e.g. local conjunction deals just with counterfeeding on the focus, OT-CC does not provide a fully general account of such cases, etc.).

We have conducted a preliminary survey, which yielded 16 cases of counterbleeding (15 environment, 1 focus) and 16 cases of counterfeeding (8 environment and 8 focus), in 25 distinct languages from a variety of language families. This lends credence to the view that opacity is a widespread phenomenon, but we have yet to examine most of these cases in detail.

We will construct our database in the first year of the grant period. We will code it using the open source software MySQL, with the assistance of a computer science major, who we will recruit with the help of Yanlei Diao of UMass computer science. For each instance of opacity, we will provide annotated examples of the data from the language, an analysis in terms of ordered rules, sketches of any prior constraint-based analyses, references to primary sources and theoretical analyses, the morphosyntactic or prosodic domains of the processes, and any attested variation or exceptionality in the application of the processes. We will also list the evidence for the posited underlying representations (e.g. from phonemic analysis or morphological alternation). We will release this database at the end of the first year, along with a solicitation for further examples that we will circulate on Language Log, the Linguist List, on other mailing lists and on social networks. We will continue to add to and develop the database throughout the grant period, and will set it up so that it can be maintained for the foreseeable future with small time investments on the part of co-PI McCarthy and PI Pater.

5 Extensions to OT-Help

The current version of OT-Help can generate typological predictions for a given set of constraints in parallel OT, HS, and variants of each with weighted constraints (that is, for four different frameworks: serial/parallel OT/HG). To aid our investigation of opacity, we will begin by implementing extensions to OT-Help to allow for work with several variants of OT that are currently capable of dealing with opacity, focusing on those that are compatible with HS. One of these is OT-CC (OT with Candidate Chains [20, 95-97]). In unmodified HS, each step of the derivation chooses the most harmonic output candidate as the input to the next step, which usually results in a single derivation (cf. [98] for discussion of ties). In contrast, OT-CC generates a set of derivations (Candidate Chains) by requiring only that each step be harmonically improving, and by not imposing the stricter restriction that it be optimal. The resulting candidate set of derivations is evaluated by standard OT markedness and faithfulness constraints, as well as by PREC constraints that penalize particular orderings of operations in the derivations. As mentioned in Section 1, we hope to improve on OT-CC in the upcoming grant period; this implementation will allow for a full assessment of both its successes and its flaws.

Our main aim is to maintain the single derivations of HS and standard generative phonology, rather than generating and evaluating sets of derivations as in OT-CC. One promising adaptation of HS is the addition of Targeted Constraints [99-103]. This appears to allow for an analysis of counterbleeding opacity of the type illustrated by Canadian raising, which as we mentioned, cannot be analyzed in plain HS [78]. Another existing possibility involves Stratal OT [19, 104-106], which is another general approach to introducing richer derivational structure to OT. Stratal OT chains together several OT grammars, usually corresponding to the Stem, Word and Phrase levels of morphosyntax. It reproduces the effects of opaque rule orderings by assigning processes to different strata and by giving the strata different constraint rankings. For example, the opaque interaction in Arabic illustrated above is the result
of stress at Stem level followed by epenthesis at Word level [106]. Stratal OT can also successfully deal with Canadian raising [107], but it has been shown to fail on a case of counterfeeding opacity in Bedouin Arabic [20]. Finally, there is also a recent modification of HS called Serial Markedness Reduction [108] that has been shown to cope successfully with the core Bedouin Arabic cases of counterfeeding and counterbleeding opacity analyzed in [20].

The Stratal implementation will be combined with all four of the current frameworks in OT-Help, and all of the HS elaborations, including the ones we develop in the grant period, will be available with weighted constraints. This means that this project will make a large number of partially distinct grammatical frameworks available for exploration in OT-Help by members of our research group and others.

6 Planned learning research: Derivational Maximum Entropy grammar

In the last grant period, we developed a general approach to hidden structure learning that built on results from an MA thesis supervised by consultant Johnson [109], and we also successfully met the main challenge in applying it to learning of HS [27]. This approach is cast in a MaxEnt version of probabilistic HG [29, 30, 110], which is popular for its usefulness in the analysis of cases of phonological variation [58, 111] and which has also been productively extended to the learning of phonotactics [31, 45, 112-114].

To show how a MaxEnt version of HS works, we'll use the opaque stress-epenthesis interaction discussed above. In the following tableaux, the constraint weights are provided underneath the constraint names, and the column labeled Harmony shows the weighted sum of each candidate’s violations. A candidates’ probability is proportional to the exponential of its Harmony. The constraint weights are as would be expected from the OT analysis. HEAD(word) has a much greater weight than PARSE(seg), which leads to probability near 1 for stress assignment in the first pass through the grammar.

**First pass**

<table>
<thead>
<tr>
<th>/katab-l-ha/</th>
<th>HEAD(word)</th>
<th>PARSE(seg)</th>
<th>DEP</th>
<th>Harmony</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. → ka’tab/ha</td>
<td>20.14</td>
<td>*</td>
<td>0</td>
<td>−9.73</td>
<td>1</td>
</tr>
<tr>
<td>b. katabilhha</td>
<td>*</td>
<td>*</td>
<td>0</td>
<td>−20.14</td>
<td>0</td>
</tr>
<tr>
<td>c. katab/ha</td>
<td>*</td>
<td>*</td>
<td>0</td>
<td>−29.87</td>
<td>0</td>
</tr>
</tbody>
</table>

PARSE(seg) has itself a much greater weight than DEP, which leads to probability near 1 for epenthesis in the second step.

**Second pass**

<table>
<thead>
<tr>
<th>ka’tab/ha</th>
<th>HEAD(word)</th>
<th>PARSE(seg)</th>
<th>DEP</th>
<th>Harmony</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. → ka’tabilha</td>
<td>20.14</td>
<td>*</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>b. ka’tab/ha</td>
<td>*</td>
<td>−9.73</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Since the probability of the indicated winner at each step is near 1, the joint probability of both steps is also near 1.
Learning a MaxEnt grammar involves finding weights that minimize the difference between the probability distribution produced by grammar and the one observed in the learning data. Such weights can be found using a variety of optimization algorithms. The weights for this stress-epenthesis analysis were found by minimizing Kullback-Leibler (KL) divergence [115], using the L-BFGS-B method [116] as implemented in R [117], with a zero minimum on weights, and a weak regularization term bounding the problem from above.

Our general approach to hidden structure learning involves summing over all of the full structures that correspond to given observed form. For example, observed [kaˈtablha] can be produced by two derivations with different orders of stress and epenthesis: /katab-l-ha/ → katabilha and /katab-l-ha/ → kaˈtablha → [kaˈtablha]. The probability that the grammar predicts for [kaˈtablha] is the sum of their probabilities. Following this summation, error minimization can proceed as just described.

The main challenge for MaxEnt learning of HS arises when we include non-monotonic derivations, which both add and remove the same structure (e.g. inserting and deleting vowels). In a probabilistic version of HS the length of these derivations is unbounded, and the set of derivations thus becomes infinite. We have shown how this challenge can be met using a Markov chain method for representing derivational probabilities [27], and we have implemented this method as an R script, which makes use of the popular statistical software's built-in optimization algorithms. We have applied this approach to learning HS for a few simple test cases, like the stress-epenthesis example above. Our proposed project moves on to broader, and more realistic, applications.

The first step of our planned learning research will be to assess our MaxEnt HS learner in comparison with other approaches to the learning of hidden structure. HS derivations require hidden structure beyond that required in parallel OT, but we do not yet know what kind of cost this adds in terms of learning. To find out, we will compare the learning of systems of word stress with differing assumptions about the structure of the grammar, and with several learning algorithms. This builds on research initiated by [7], who proposed a method for learning hidden structure in OT, and tested it using a set of 124 metrical systems. These initial results have been improved upon by using probabilistic versions of OT and HG [55, 118, 119] (see also [120]). This test set has provided a useful basis for theory comparison, but there are two problems with it. One is that its developers do not state whether all the patterns exist in natural languages, and our comparison of some them to the typologically attested patterns of which we are aware leads us to suspect that at least some do not. The other problem is that there seems to be no principled basis for the choice of this particular set – its members do not exhaust the possibilities in terms of what the metrical system can express.

We will thus construct a new set of test cases. To do so, we will start with 109 cross-linguistically attested stress patterns, which can be generated by freely available web-based software [121]. We will use OT-Help to determine which of those languages can be captured by the HS stress system whose development was supported by our last project [49], and by two comparable parallel OT stress systems whose typological predictions have been well-studied [122]. To test the relative learnability of HS as compared with parallel OT, we will examine how well our general approach to learning hidden structure does with each of these three grammatical systems on the languages they can all represent. To see how well our MaxEnt learner does in comparison with previous approaches to learning hidden structure, we will compare its performance on the parallel OT systems to implementations of earlier proposals [55, 119].
In preliminary research using the currently available metrical phonology test set, we have found that when we start our learner with zero weights, and use a weak regularization term, its performance is similar to that achieved in earlier work with other versions of probabilistic HG [55], but not as good as the current best results [119]. We will test several methods used in machine learning for avoiding the traps (local minima) that can result from structural ambiguity in MaxEnt learning. One simple approach, which did yield some improvement in our pilot work, is to run the learner with some number of random initializations of the weights. Another approach is use a stronger regularization term [123], and pilot tests also show this improves performance on these problems. Finally, the literature contains various applicable forms of simulated annealing [124, 125] that essentially introduce some degree of gradually reduced randomness to the algorithm, which helps to avoid hidden structure traps.

The second stage of our learning research will be to study the learning of opacity, with the goal of comparing different grammatical theories in terms of how they contribute to the success of learning cross-linguistically attested patterns. Another related focus of our research on hidden structure learning in the last grant period was the learning of underlying representations (URs). Learning of opacity requires not only the learning of hidden intermediate steps, but also the learning of URs that differ from those observed on the surface. The feasibility of learning such “abstract” underlying representations has long been controversial [126, 127]. In our earlier research [128], we studied cases of UR learning in which the space of possible URs for a morpheme can be limited to its surface-observed alternants, showing that this covers a wider range of data than usually assumed. In the upcoming grant period we will extend this approach to the more abstract URs needed for opacity.

Our account of UR learning assumes that the input to the phonological component is a meaning, or a set of morpho-syntactic features, and that each possible UR for each morpheme is demanded by a separate constraint [109, 129-131]. Learning URs is the induction of these UR constraints, and the setting of their weights. Though we have not yet implemented an induction procedure for UR constraints, we can illustrate this approach for Canadian English ‘mitre’ [mairə] by providing the learner with two of these constraints: ‘mitre’→/maitə/ and ‘mitre’→/maidə/ . Each of these is violated if the UR differs from the specified one, which has either /t/ or /d/ for the flap. This illustration, which has a three-step derivation, involves learning in a MaxEnt version of Stratal OT. The UR is chosen in the first step, and the relatively high weight on the constraint demanding /maitə/ leads to it having probability near 1:

\[
\begin{array}{|c|c|c|c|c|}
\hline
UR choice & ‘mitre’→/maitə/ & ‘mitre’→/maidə/ & Harmony & p \\
\hline
a. & /maitə/ & & 13.12 & 0 & 1 \\
b. & /maidə/ & * & & \hline
\end{array}
\]

The second and third steps are the applications of Stem and Phrase level grammars. These include a Markedness constraint that bans the unraised diphthong before a voiceless consonant, *AIt, and one that bans an intervocalic alveolar stop, FlAP, as well as conflicting faithfulness IDENT-V and IDENT-C respectively. In the Stem level, the high weight of *AIt and zero weight of IDENT-V lead to extremely high probability of raising. Flapping is blocked because of the high weight of IDENT-C relative to FlAP.
Unlike the HS derivations shown above in which the ranks or weights of the constraints remain constant, in Stratal OT or HG the grammar can change across levels. In this derivation, at the Phrase level the relationship between the Markedness and Fatifulness constraints is reversed from the Stem level, resulting in flapping, and preservation of the raised vowel.

This example shows how cases of opacity can require the consideration of non-observed alternants as URs: the underlying /t/ that is required for the conditioning of the raised [ʌɪ] is a fixed surface [ɾ] in ‘mitre’.

For this example, and many others like it, one possibility for constraint induction is that the learner expands the set of UR constraints by taking a ‘free ride’ on the observation of alternations in other morphemes: observed [t]/[ɾ] and [d]/[ɾ] alternations lead to the postulation of UR constraints demanding underlying /t/ and /d/ for any surface [ɾ]. An implementation of such a free ride algorithm for UR constraint induction will be our first step in studying the feasibility of the learning of abstract URs in this framework.

The weights for the stratal MaxEnt Canadian raising analysis were found using the R-implemented method described for the stress-epenthesis case. The learner was provided with just the meaning ‘mitre’ and the surface form [mʌɪɾɚ], and found the constraint weights shown above. The learner was also provided with other forms in the language, which demonstrate transparent raising at the Stem level (‘lifer’ [lʌɪɾɚ]), absence of raising at the Phrase level (‘lie for’ [laɪɾə]), and presence of flapping at the Phrase level (‘lie to’ [laɪɾ]). This illustrates the flexibility of our learning model; it works not only for HS, but also for stratal OT, and we see no barriers to extending it to HS with Targeted Constraints. OT-CC appears to be more of a

\[\text{Stem level}\]

<table>
<thead>
<tr>
<th></th>
<th>*AIT</th>
<th>IDENT-V</th>
<th>FLAP</th>
<th>IDENT-C</th>
<th>Harmony</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>14.52</td>
<td>0</td>
<td></td>
<td>13.83</td>
<td>−13.83</td>
<td>0</td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−14.52</td>
<td>0</td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−13.83</td>
<td>0</td>
</tr>
</tbody>
</table>

\[\text{Phrase level}\]

<table>
<thead>
<tr>
<th></th>
<th>*AIT</th>
<th>IDENT-V</th>
<th>FLAP</th>
<th>IDENT-C</th>
<th>Harmony</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td>14.91</td>
<td></td>
<td></td>
<td>−14.23</td>
<td>0</td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−29.14</td>
<td>0</td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−14.91</td>
<td>0</td>
</tr>
</tbody>
</table>

\[\text{2 It is of course possible that human learners are exposed to slow speech variants with unflapped /t/. It does not appear that all instances of abstract URs needed for opaque analyses are seen in slow speech variants, or in other variation – though whether the availability of such evidence is connected to ease of learning by computers or humans, and to the productivity and diachronic persistence of opaque alternations is worthy of further study (see relatedly [145]).}\]
challenge due to its global PREC constraints. The challenge that OT-CC poses for our approach to learning forms part of the motivation for our seeking alternatives within the single-derivation approach to HS.

We will compare Stratal OT, HS with Targeted Constraints, and further HS approaches for their success in learning opacity in our MaxEnt hidden structure learning framework. Based on study of the learning of metrical phonology, we will select the best-performing elaborations of our learning model that involve each of: (1) Repeated random initialization (2) Strong regularization (decreased over a training schedule) and (3) Simulated annealing. It is important to examine a set of algorithms, because it is possible that the learning spaces introduced by each of the theories of opacity might be better navigated by one or another approach.

Drawing on our work on the opacity database, we will select a representative set of patterns that differ along several dimensions, such as counterbleeding vs. counterfeeding, focus vs. environment, and morphological scope of the rules (which is especially important to learning in Stratal OT [19, 107]). We will aim to make this set cover the full range of opaque patterns found cross-linguistically. We will also include representative cases of variation and exceptionality, which can be handled by our general grammar and learning framework [128]. We will then systematically compare the success in learning for each theory that can handle some type of pattern, against other theories that handle that same one.

Given the structure of our objective function, which sums over hidden structures that correspond to a single overt form, an appropriate measure of success is whether the overt form in a given tableau has higher probability than any other overt form [128]. We will measure the overall success of each combination of grammar and learning algorithm over a set of languages in two ways. First, we will count the number of languages on which all of the overt forms are successfully learned, as just defined. Second, we will sum over all of the overt forms over all of the languages. These measures will constitute our means of determining the relative learnability of the grammatical theories under our assumptions about learning.

The most straightforward interpretation of our results will be that the grammatical theory with the highest success rate in learning is to be preferred, modulo any differences in typological fitness uncovered in the research described in Section 3. Since these patterns do seem to be acquired by human learners, it is a reasonable goal for a theory of learning to be able to find a correct grammar for all of them. It has in fact long been speculated that opacity is only marginally learnable, in that certain forms of it tend to be eliminated in language change [133-135]. Ultimately, therefore, we would want a theory of learning and grammar that also reflects the relative difficulty of learning different types of opaque patterns. We will examine the results of our learning simulations from this perspective as well: we will identify instances of opacity that pose particularly severe learning problems and compare these to the cases that have been identified thus far as potentially difficult to learn.

As well as meeting independently important theoretical goals, we see this research on the representation and learning of opacity as forming the foundation for future studies of human learning of these kinds of phonological processes. In particular, we expect that the hypotheses that emerge from this study about the relative ease of learning different types of opaque pattern will be amenable to testing using the increasingly popular method of artificial language learning (reviewed in [113]). The MaxEnt learning model that we will be developing can not only be used to assess the learnability of patterns, but it can also be used to generate predicted learning trajectories when gradient descent is used to find the weights [110, 114]. Recent collaborative work on artificial language learning involving PI Pater has found support for the predictions of a
MaxEnt model of the learning of phonotactic patterns of differing structural types [114]. The proposed project will allow this work to be extended to a broader range of phonological patterns.

7 Schedule of work, management plan, and assessment

The research plan described above involves 6 interacting components:

i. Opacity database
ii. HS theory of opacity
iii. Extensions of OT-Help software
iv. Stress and opacity learning test batteries
v. R MaxEnt learning package
vi. Learnability tests of theories of grammar with MaxEnt learning models

Year 1 will focus on the creation of the opacity database, on the implementation of existing theories of opacity as extensions to OT-Help, and on stress problems for hidden structure learning. By the end of the summer following Year 1, we will publicly release the opacity database and begin to solicit additions from the phonological community. We will also release OT-Help 3.0, an R package for MaxEnt learning for parallel OT, HS, and Stratal OT, and the stress learning test battery.

Year 2 will focus on the development of alternative theories of opacity in HS, and on opacity problems for hidden structure learning. This research will make use of the database and software released in Year 1, which will lead to refinements and further documentation. In this year, we will release database and software updates and as well the opacity learning test battery.

Year 3 will focus on the comparison of the HS theories of opacity with other derivational extensions of OT, in particular stratal OT, both in terms of typological predictions and learning. This will involve implementation of the new HS theories in OT-Help and the R learning package.

Some of the preliminary research described in our proposal has been presented at local workshops and at international conferences. We are now starting to prepare it for journal submission, so even the early stages of this grant should see substantial dissemination of the research. In addition, the staging of our research over the three-year period of the grant will lead to natural sets of results for conference and journal paper submissions each year. Our target journals are typically the leading ones in theoretical linguistics (e.g. Linguistic Inquiry, NLLT, Phonology) and we also aim to reach an even broader audience (e.g. Language, Cognitive Science). We were successful in placing our work in these journals in the last grant period. We have also identified Language Acquisition, which has recently broadened its scope, as a potential home for our computational learning research, in addition to more computationally specialized journals (e.g. Proceedings of ACL/SIGMORPHON).

Throughout the course of the grant we will hold weekly lab meetings, to facilitate and assess the progress of our research. PI Pater will lead the learning research, in collaboration with the graduate student RAs, and in consultation with McCarthy, as well as Mark Johnson and Robert Staubs. Consultation with Johnson and Staubs, which will also be necessary for software development and will also involve graduate student RAs, will take place via hour-long Skype sessions, biweekly with Johnson, and weekly with Staubs. The Year 2 and 3 RAs will also consult in person with Johnson for one week each on the learning research. Co-PI McCarthy will lead the research on HS theories of opacity, collaborating with the RAs and consulting with Pater.
References

26. Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John McCarthy and Joe Pater, *OT-Help 2.0 [computer program]*. 2010, University of Massachusetts Amherst: Amherst, MA. Available at http://web.linguist.umass.edu/~OTHelp/.
32. Staubs, Robert, *HG in R [computer program]*. 2011, University of Massachusetts Amherst: Amherst, MA.


75. Elfner, Emily, *Syntax-Prosody Interactions in Irish*. 2012, University of Massachusetts Amherst: Amherst, MA.


89. Gouskova, Maria, *Deriving Economy: Syncope in Optimality Theory*. 2003, University of Massachusetts Amherst: Amherst, MA. Available on Rutgers Optimality Archive, ROA-610.


